On Minimizing the MSE in the Presence of Channel State Information Errors

Gábor Fodor*, Piergiuseppe Di Marco*, Miklós Telek†

*Ericsson Research, Stockholm, Sweden. E-mail: gabor.fodor@ericsson.com
†Royal Institute of Technology, Stockholm, Sweden. E-mail: pidm@kth.se
‡Budapest University of Technology and Economics, Budapest, Hungary. E-mail: telek@hit.bme.hu

Abstract—We consider the uplink of a multiuser multiple input multiple output (MU MIMO) system, in which the base station acquires channel state information (CSI) for which the estimation error depends on the resources assigned to the uplink pilot symbols. For this system, we first derive the receiver that minimizes the mean square error (MSE) of the uplink detected data symbols, as opposed to the naïve receiver that does not minimize the MSE in the presence of CSI errors. We then derive a closed form expression for the MSE as a function of the employed pilot-to-data power ratio, number of antennas and the MU MIMO interference power. This expression allows us to gain the insight that the gain of using the actual MMSE receiver as opposed to the naïve receiver becomes particularly important when the number of BS antennas is large.

I. INTRODUCTION

In multiuser multiple input multiple output (MU MIMO) systems, the fundamental trade-off between spending resources on channel state information (CSI) acquisition and data transmission is known to affect the performance in terms of spectral and energy efficiency [1], [2]. Therefore, balancing the pilot-to-data power ratio (PDPR) [3] and determining the number of pilot and data symbols are important aspects of designing MIMO systems [4], [5], [6]. From a different perspective, a related work combined a transmitter employing a linear dispersion code (LDC) and a linear minimum mean square error (MMSE) detector at the receiver [7]. It has been found that optimizing the average normalized mean square error (MSE) is relevant for detectors employing a linear front end and helps designing optimal transmit strategies. In this paper we consider the uplink of a MIMO system employing an MMSE receiver for data reception [8]. The MMSE receiver is initialized by the estimates of the channel state information rather than assuming the availability of perfect CSI. Thus, our contribution to the existing literature is two-fold:

1) We derive the actual MMSE receiver that, – in contrast to the classical or naïve formula [9] – minimizes the MSE of the estimated uplink data symbols in the presence of PDPR dependent estimation errors.

2) Secondly, we derive a closed form exact expression for the MSE, as a function of not only the PDPR but also the number of antennas. This exact formula allows us to arrive at the key insight that employing the actual MMSE gives large gains as the number of antennas grows large.

II. SYSTEM AND CHANNEL ESTIMATION MODEL

We consider the uplink of a MU MIMO system, in which the mobile stations (MS) transmit orthogonal pilot sequences \( s = [s_1, ..., s_{\tau_p}]^T \in \mathbb{C}^{\tau_p \times 1} \), in which each pilot symbol is scaled as \(|s_i|^2 = 1\), for \( i = 1, ..., \tau_p \). The pilot sequences are constructed such that they remain orthogonal as long as the number of spatially multiplexed users is maximum \( \tau_p \). Specifically, without loss of generality, we assume that the number MU-MIMO users is \( K \leq \tau_p \). In practice, \( K \ll N_r \), where \( N_r \) is the number of antennas at the BS.

In this paper we assume a comb type arrangement of the pilot symbols [10]. Given \( F \) subcarriers in the coherence bandwidth, a fraction of \( \tau_p \) subcarriers are allocated to the pilot and \( F_d = F - \tau_p \) subcarriers are allocated to the data symbols. Each MS transmits at a constant power \( P_{tot} \), however, the transmission power can be distributed unequally in each subcarrier. In particular, considering a transmitted power \( P_p \) for each pilot symbol and \( P \) for each data symbol transmission, the sum constraint of \( \tau_p P_p + (F - \tau_p) P = P_{tot} \) is enforced. Thus, the \( N_r \times \tau_p \) matrix of the received pilot signal from a specific MS at the BS can be conveniently written as:

\[
Y^p = \alpha \sqrt{P_p} h s^T + N, \tag{1}
\]

where we assume that \( h \in \mathbb{C}^{N_r \times 1} \) is a circular symmetric complex normal distributed column vector with mean vector 0 and covariance matrix \( C \) (of size \( N_r \)), denoted as \( h \sim \mathcal{CN}(0, C) \), \( \alpha \) accounts for the propagation loss, \( N \in \mathbb{C}^{N_r \times \tau_p} \) is the spatially and temporally additive white Gaussian noise (AWGN) with element-wise variance \( \sigma^2 \), where the index \( p \) refers to the noise power on the received pilot signal.

In this paper we assume that the BS uses the popular least square (LS) estimator that relies on correlating the received signal with the known pilot sequence. Note that our methodology to determine the MSE of the received data is not confined to the LS estimator, but is directly applicable to an MMSE or other channel estimation techniques as well. For each MS, the BS utilizes pilot sequence orthogonality and estimates the channel based on (1) assuming:

\[
\hat{h} = h + w = \frac{1}{\alpha \sqrt{P_p}} Y^p s^*(s^T s)^{-1} = h + \frac{1}{\alpha \sqrt{P_p \tau_p}} N s^*, \tag{2}
\]

where \( s^* = [s_1^*, ..., s_{\tau_p}^*]^T \in \mathbb{C}^{\tau_p \times 1} \) denotes the vector of pilot symbols and \( (s^T s)^{-1} = \tau_p \). By considering \( h \sim \mathcal{CN}(0, C) \),
it follows that the estimated channel \( \hat{h} \) is a circular symmetric complex normal distributed vector \( \hat{h} \sim CN(0, R) \), with
\[
R = E(\hat{h} \hat{h}^H) = C + \frac{\sigma_d^2}{\alpha^2 P r_p} I_{N_r}.
\]
As it was shown in [10], the distribution of the channel realization \( h \) conditioned on the estimate \( \hat{h} \) is normally distributed as follows:
\[
(h \mid \hat{h}) \sim \mathcal{D} \hat{h} + \mathcal{C}N(0, Q),
\]
where \( D \triangleq CR^{-1} \) and \( Q \triangleq C - CR^{-1} C. \)

III. LINEAR MMSE RECEIVER

A. Received Data Signal Model

The MU-MIMO received data signal at the BS can be written as:
\[
y = \alpha_k h_k \sqrt{P_k} x_k + \sum_{\substack{k' \neq k \\text{Other users}}} \alpha_k h_k \sqrt{P_k} x_k + n_d,
\]
where \( \alpha_k \cdot h_k \) is the \( M \times 1 \) vector channel including large and small scale fading between User-\( k \) and the BS (\( \alpha_k \) and \( h_k \) respectively). \( x_k \) is the transmitted data symbol by User-\( k \) and \( n_d \) emphasizes the noise on the received signal.

B. Employing an MMSE Receiver at the BS

In this paper the BS employs an MMSE receiver \( G_{k} \in \mathbb{C}^{1 \times N_r} \) to estimate the data symbol transmitted by User-\( k \). We recall that the MMSE receiver aims at minimizing the mean-square error between the estimate \( \hat{G}_{k}y \) and the transmitted symbol \( x_k \):
\[
\hat{G}_{k} = \arg \min_{\hat{G}} E[\| \hat{G} y - x_k \|^2].
\]
When the BS employs a naïve receiver, the estimated channel is taken as if it was the actual channel:
\[
\hat{G}_{k}^{\text{naive}} = \alpha_k \sqrt{P_k} h_k^H (\alpha_k^2 P_k h_k h_k^H + \sigma_d^2 I)^{-1}.
\]
As we shall see, this receiver does not minimize the MSE.

C. Determining the Actual MMSE Receiver Matrix

In this section we determine the MMSE receiver matrix \( G_{k} \) that the BS should use to demodulate the received data signal such that the data estimation error for User-\( k \) is minimized taking explicitly account that the BS has access only to the estimated channels \( \hat{h}_{k} \), as opposed to the naïve receiver that minimizes the MSE only when perfect channel estimation is assumed. To this end, we consider the MSE of the estimated data symbols of the tagged User-\( k \); obtained from the signal model of (5) using a receiver vector \( G_{k} \):

\[
\text{MSE}(G_{k}, h_1, \ldots, h_K) = E_{x_k, n_d} \{ \| G_{k} y - x_k \|^2 \} =
\]
\[
= E_{x_k, n_d} \left[ (G_{k} \alpha_k h_k \sqrt{P_k} - 1) x_k + \sum_{k' \neq k} G_{k} \alpha_k h_k \sqrt{P_k} x_k + G_{k} n_d \right]^2
\]
\[
= E_{x_k, n_d} \left[ (G_{k} \alpha_k h_k \sqrt{P_k} - 1) x_k^2 + \sum_{k' \neq k} P_k E_{x_k, n_d} \{ G_{k} \alpha_k h_k x_k \}^2 \right]
\]
\[
+ E_{x_k, n_d} \{ G_{k} n_d \}^2,
\]
where we utilized that \( E\{x_k\} = 0 \) and \( E\{n_d\} = 0 \). Additionally, utilizing \( E\{x_k x_k^H\} = 1 \) and \( E\{n_d n_d^H\} = \sigma_d^2 I_{N_d} \), we have:
\[
\text{MSE}(G_{k}, h_1, \ldots, h_K) = \left| G_{k} \alpha_k h_k \sqrt{P_k} - 1 \right|^2 + \sum_{k' \neq k} \left| P_k G_{k} \alpha_k \right|^2 + \sigma_d^2 G_{k} G_{k}^H,
\]
from which our first result follows.

Result 1. When the BS uses the receiver vector \( G_{k} \), the MSE of the received data symbols of the tagged user \( k \) assuming perfect channel state information at the base station is:
\[
\text{MSE}(G_{k}, h_1, \ldots, h_K) = \left| G_{k} \alpha_k h_k \sqrt{P_k} - 1 \right|^2 + \sum_{k' \neq k} \left| P_k G_{k} \alpha_k \right|^2 + \sigma_d^2 G_{k} G_{k}^H + 1
\]
\[+ \sigma_d^2 G_{k} G_{k}^H + \sum_{k' \neq k} \alpha_k^2 P_k G_{k} C_{k} G_{k}^H.
\]

Although this result is useful, we need an expression for the MSE as a function of \( \hat{h} \), rather than \( h \).

Result 2. The MSE of the received data symbols of the tagged user \( k \) as a function of the estimated channel at the BS is:
\[
\text{MSE}(\hat{G}_{k}, \hat{h}_1, \ldots, \hat{h}_K) = \text{MSE}(G_{k}, h_1, \ldots, h_K)
\]
\[= \alpha_k^2 P_k G_{k} (D_{k} \hat{h}_k \hat{h}_k^H + Q_{k}) G_{k}^H + \sigma_d^2 G_{k} G_{k}^H + 1
\]
\[+ \sum_{k' \neq k} \alpha_k^2 P_k G_{k} C_{k} G_{k}^H - \alpha_k \sqrt{P_k} (G_{k} D_{k} \hat{h}_k + \hat{h}_k D_{k}^H G_{k}^H).
\]

Using these results, we are in the position of deriving the optimal MU MIMO receiver vector for User-\( k \):

Proposition 3. The optimal \( G_{k}^* \) can be derived as:
\[
G_{k}^* = \alpha_k \sqrt{P_k} \hat{h}_k \hat{h}_k^H D_{k}^{-1}
\]
\[= \left( \alpha_k^2 P_k (D_{k} \hat{h}_k \hat{h}_k^H + Q_{k}) + \sum_{k' \neq k} \alpha_k^2 P_k C_{k} + \sigma_d^2 I \right)^{-1}.
\]

IV. DETERMINING THE MSE OF THE RECEIVED DATA SYMBOLS WITH OPTIMAL \( G^* \)

In the case of proper antenna spacing, the channel covariance matrices can be modeled as \( C_{k} = c_k I \), which for \( k = \kappa \) implies \( D_{k} = d_k I \), \( Q_{k} = q_k I \). In this case, the MSE as a function of the estimated channel can be obtained as follows.

Lemma 4. In the case of uncorrelated antennas at the BS, when the BS employs the optimal receiver \( G_{k}^* \), the mean square error of the received data symbols can be expressed as:
\[
\text{MSE}(\hat{h}_k) = -2\alpha_k \sqrt{P_k} d_k \| \hat{h}_k \|^2 + 1
\]
\[+ g_k^2 \cdot \left( \alpha_k^2 P_k d_k^2 \| \hat{h}_k \|^2 + \left( \alpha_k^2 P_k q_k + \sum_{k' \neq k} \alpha_k^2 P_k c_k + \sigma_d^2 \right) \| \hat{h}_k \|^2 \right)
\]
where
\[
g_k \triangleq \frac{\alpha_k \sqrt{P_k} d_k}{\alpha_k \sqrt{P_k} d_k^2 + q_k} + \sum_{k' \neq k} \alpha_k^2 P_k c_k + \sigma_d^2.
\]
We can now derive the unconditional MSE from \( \text{MSE} = \mathbb{E}_{\mathbf{h}_n} \text{MSE}(\hat{\mathbf{h}}_n) \) based on the distribution of \( \hat{\mathbf{h}}_n \), which we recall from (3) as \( \hat{\mathbf{h}}_n \sim \mathcal{CN}(0, \mathbf{R}_n) \).

V. CALCULATING THE UNCONDITIONAL MSE

To calculate the unconditional MSE, notice that the MSE(\( \hat{\mathbf{h}}_n \)) depends on \( \mathbf{h}_n \) only through \( ||\hat{\mathbf{h}}_n||^2 \). Thus, we can conveniently introduce \( Y_n \triangleq ||\hat{\mathbf{h}}_n||^2 \), substitute \( g_n \) into (13) and, by inspecting (13), introduce the following notations:

\[
T_1 \triangleq g_n^2 \left( \alpha^2 b_p + d_p^2 ||\hat{\mathbf{h}}_n||^4 \right) = \frac{\alpha^2 Y_n^2}{(b_n + s_n Y_n)^2},
\]

where we introduced the notation \( s_n \triangleq d_p^2 b_p \), \( p_n \triangleq \alpha^2 b_p \), \( \sigma_n^2 \triangleq \sum_{k \neq n} \alpha_k^2 b_k c_k + \sigma_d^2 \) and \( Y_n \triangleq ||\hat{\mathbf{h}}_n||^2 \) and \( b_n \triangleq q_n p_n + \sigma_n^2 r_n \). Similarly:

\[
T_2 \triangleq g_n^2 \left( \alpha^2 b_p + \sum_{k \neq n} \alpha_k^2 b_k c_k + \sigma_d^2 \right) ||\hat{\mathbf{h}}_n||^2 = \frac{b_n s_n Y_n}{(b_n + s_n Y_n)^2},
\]

\[
T_3 \triangleq 2d_n \alpha \sqrt{P_{tot}} ||\hat{\mathbf{h}}_n|| \cdot g_n = \frac{2s_n Y_n}{b_n + s_n Y_n}.
\]

Thus, we can prove the following proposition, which will serve as the basis for numerical evaluations.

**Proposition 5.** The unconditional MSE of the received data symbols of User-\( \kappa \) when the BS uses the optimal \( G_\kappa \) receiver is as follows.

\[
\text{MSE} = \frac{b_n}{s_n} \left( e^{\frac{b_n}{s_n}} - \frac{b_n}{s_n} \right) E_{in} \left( N_r, \frac{b_n}{s_n} \right) - s_n r
\]

\[
+ \frac{N_r}{s_n} \left( e^{\frac{b_n}{s_n}} (b_n + (1 + N_r) s_n r) E_{in} \left( 1 + N_r, \frac{b_n}{s_n} \right) - s_n r \right)
\]

\[
- 2 \cdot e^{\frac{b_n}{s_n}} N_r E_{in} \left( 1 + N_r, \frac{b_n}{s_n} \right) + 1,
\]

where \( E_{in}(n, z) \triangleq \int_1^{\infty} e^{-zt} t^n \, dt \) is a standard exponential integral function.

VI. NUMERICAL RESULTS AND CONCLUDING REMARKS

| Table I: System Parameters |

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of antennas</td>
<td>( N_r = 2, 4, 8, 10, 20, 50, 100, 500 )</td>
</tr>
<tr>
<td>Path Loss of Tagged User-( \kappa )</td>
<td>( \alpha = 40, 45, 50 ) dB</td>
</tr>
<tr>
<td>Number of pilot and data symbols</td>
<td>( \tau_p = 1; \tau_d = 11 )</td>
</tr>
<tr>
<td>Power budget</td>
<td>( \tau_p P_p + \tau_d P = P_{tot} = 250 ) mW</td>
</tr>
</tbody>
</table>

In this section we consider a single cell single user MIMO system, in which the mobile terminal is equipped with a single transmit antenna, whereas the BS employs \( N_r \) receive antennas. Note that the performance characteristics of the proposed MMSE receiver as compared with the naïve receiver are similar in the multi-user MIMO case from the perspective of the tagged user, since the proposed receiver treats the multi-user interference as noise according to (12). The key input parameters to this system that are necessary to obtain numerical results using the MSE derivation in this paper (ultimately relying on Proposition 5) are listed in Table I.

![Figure 1: MSE as the function of the pilot power \( P_p \) assuming a fixed pilot-data power budget with \( N_r = 20 \) and \( N_r = 500 \) number of antennas when using the naïve receiver and the MMSE receiver.](image)

Figure 1 compares the performance of the system in which the number of antennas at the BS grows large (\( N_r = 500 \)). As expected, given a fixed sum power budget of \( \tau_p P_p + \tau_d P = P_{tot} = 250 \) mW, the optimal pilot-data power allocation becomes non trivial as it depends on the number of antennas, path loss and the receiver structure. The minimum value of the MSE in all cases are marked with a dot, which clearly indicates that the achievable minimum MSE with this power budget is significantly lower when employing the MMSE receiver.

Figure 2 shows the achievable minimum MSE value and the optimal pilot power setting as the function of the number of antennas at the base station. First, notice that the gain in terms of achievable minimum MSE increases as the number of antennas increases.

For example, at \( N_r = 500 \) the gain is around 6 dB. Interestingly, the pilot power setting that minimizes the MSE does not depend on the number of antennas when using the MMSE receiver, whereas it increases with the number of antennas in the case of the naïve receiver. The intuitive explanation for this is that in the case of uncorrelated antennas, according to equation (3), the diagonal elements of the covariance of the CSI error does not depend on the number of antennas, although the size of the matrix does. Thus, the pilot-data ration when using the MMSE receiver does not depend on the number of antennas, as opposed to the naïve receiver case, which does not minimize the MSE. The formal proof of this phenomenon is planned for future work.

APPENDIX I: PROOF OF RESULT 1

From (9) it follows, that focusing on the tagged User-\( \kappa \):

\[
\text{MSE}(G_\kappa, h_\kappa) = \mathbb{E}_{h_1, \ldots, h_{\kappa-1}, h_{\kappa+1}, \ldots, h_K} |G_\kappa, h_\kappa|^2 + \sum_{k, k \neq \kappa} \alpha_k^2 P_k E_{h_k} |G_\kappa, h_k|^2 + \sigma_d^2 G_\kappa G_\kappa^H.
\]
Recognizing that [10]
\[ |G_\alpha \sqrt{F} - 1|^2 = \alpha^2 P G h^H G H - \alpha \sqrt{P (G h h^H G H)} + 1, \]
eq \frac{P_\kappa}{\sum_{k \neq \kappa} P_{\kappa} c_k + \sigma_d^2} \hat{h}_k^H
\]
and
\[ e_{h_k} |G_h h_k|^2 = G_h e_{h_k} |h_k|^2 G_h^H = G_h C_h G_h^H, \]
the result follows.

**APPENDIX II: PROOF OF RESULT 2**

Utilizing \((h_k | \hat{h}_k) \sim D_h \hat{h}_k + CN(0, Q_h), \) where \(D_h = C_h R_h^{-1}, R_h = C_h + C_h^w \text{ and } Q_h = C_h - C_h R_h^{-1} C_h, \)
and, by averaging over \(h_k | \hat{h}_k, \) and following the technique proposed in [10], the result follows.

**APPENDIX III: PROOF OF PROPOSITION 3**

To derive the optimal \(G_\kappa, \) we rewrite MSE \((G_\kappa, \hat{h}_k) \) in quadratic form of \((x A x^H - x B - B^H x^H + 1):\)

**APPENDIX IV: PROOF OF LEMMA 4**

If \(C_\kappa = c_\kappa I, \) implying \(D_\kappa = d_\kappa I, \) \(Q_\kappa = q_\kappa I \) and the optimal \(G_\kappa^* \) can be written as:

\[ G_\kappa^* = \frac{\alpha_\kappa \sqrt{P_{\kappa}} d_\kappa}{\alpha_\kappa^2 P_{\kappa} (d_\kappa^2 |\hat{h}_k|^2 + q_\kappa) + \sum_{i \neq \kappa} \alpha_i^2 P_i c_i + \sigma_d^2 I} \hat{h}_k^H \]
\[ \triangleq g_\kappa \cdot \hat{h}_k^H. \] (15)

Substituting \(G_\kappa^* \) into the MSE of Result 2 gives the lemma.

**APPENDIX V: PROOF OF PROPOSITION 5**

Recognizing that \(Y_\kappa \) is Gamma distributed, the density function of \(Y_\kappa \) for convenience:

\[ f_Y (x) = \frac{r^{-N_\kappa} x^{N_\kappa-1} e^{-x/r}}{(N_\kappa - 1)!} \quad x > 0. \] (16)

Proposition (5) follows from Lemma (4) taking the average of MSE \((\hat{h}_k) \) using the following integrals:

\[ \int_{x=0}^{\infty} T_1 f_{Y_\kappa} (x) dx = \frac{N_\kappa}{(N_\kappa - 1)!} \cdot \left( b_\kappa + (1 + N_\kappa) s_\kappa r \right) E_{in} \left( 1 + N_\kappa \cdot \frac{b_\kappa}{s_\kappa r} \right) - s_\kappa r, \]

\[ \int_{x=0}^{\infty} T_2 f_{Y_\kappa} (x) dx = \frac{b_\kappa}{s_\kappa r} \left( b_\kappa + N_\kappa s_\kappa r \right) E_{in} \left( N_\kappa \cdot \frac{b_\kappa}{s_\kappa r} \right) - s_\kappa r, \]

\[ \int_{x=0}^{\infty} T_3 f_{Y_\kappa} (x) dx = 2 \cdot \frac{b_\kappa}{s_\kappa r} N_\kappa E_{in} \left( 1 + N_\kappa \cdot \frac{b_\kappa}{s_\kappa r} \right). \]

**REFERENCES**


